Available online at www.sciencedirect.com

ScienceDirect

Genomics 89 (2007) 378–384

GENOMICS

www.elsevier.com/locate/ygeno

Promoter profiling and coexpression data analysis identifies 24 novel genes that are coregulated with AMPA receptor genes, GRIAs

Allen Chong [a,*], Zhuo Zhang [a], Kwok Pui Choi [b,c], Vidhu Choudhary [d], Mustafa B.A. Djamgoz [e], Guanglan Zhang [a], Vladimir B. Bajic [a,d]

[a] *Molecular Bioinformatics Group, Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, Singapore*
[b] *Department of Mathematics, National University of Singapore, Singapore 119613, Singapore*
[c] *Department of Statistics and Applied Probability, National University of Singapore, Singapore 119613, Singapore*
[d] *South African National Bioinformatics Institute, University of the Western Cape, Bellville 7535, South Africa*
[e] *Division of Cell and Molecular Biology, Sir Alexander Fleming Building, Imperial College London, South Kensington Campus, London SW7 2AZ, UK*

Received 13 April 2006; accepted 24 November 2006

Abstract

We identified a set of transcriptional elements that are conserved and overrepresented within the promoters of human, mouse, and rat GRIAs by comparing these promoters against a collection of 10,741 gene promoters. Cells regulate functional groups of genes by coordinating the transcriptional and/or posttranscriptional mRNA levels of interacting genes. As such, it is expected that functional groups of genes share the same transcriptional features within their promoters. We found 47 genes whose promoters contain the same combination of transcriptional elements that are overrepresented within the promoters of the GRIA gene family. Coexpressed genes may be transcriptionally coregulated, which in turn suggests that these genes may play complementary roles within a particular functional context. Using microarray expression data, we found 24 (of the 47) genes that share not only a similar promoter profile with GRIAs but also a well-correlated gene expression profile and, thus, we believe these to be coregulated with GRIAs.

Keywords: Ionotropic AMPA glutamate receptor (GRIA/GluR1-4); Transcription factor binding sites (CDP CR1, Sp3, BACH2, GKLF, PU.1, MZF1, GATA2, STAT1, STAT3, STAT6, ELF1, Pax4); Gene expression regulation (KLK6, CLSTN3, RAB3B, IL16, BAI2, FAM107A, KLHL18, BEX1, DSN1, PIGC, DKKL1, VAMP4, NFKBIB, CD1C, FLT3LG, CAPN11, ARMC8, SETD4, CASZ1, FLJ10847, ADAMTS9, ARL6IP2, ALDH8A1, WDR71); Microarray; Mammal; Genome; Computational biology

The AMPA receptor (GRIA) mediates fast excitatory synaptic transmission. Its function is essential for diverse brain activities. For example, AMPA receptors are implicated in hippocampal long-term potentiation, the processes of learning and memory acquisition, brain and spinal injury, and seizure as a result of ischemia, hypoxia, or physical trauma [1,2]. As such, the study of the mechanisms controlling their expression is of much interest.

AMPA receptors are heterooligomeric molecules composed of GRIA1 to GRIA4 subunits (using HGNC nomenclature; also known as GluR1-4 or GluRA-D). The expression of AMPA receptor subunits is differentially regulated during development and can be affected by extraneous conditions, for example, ischemia. A downregulation of GRIA2 gene expression in the hippocampal CA1 neurons is observed following global ischemia, leading to an increased Ca^{2+} influx through AMPA receptors in response to endogenous glutamate and eventually to neuronal death [3]. Conversely, increased GRIA2 subunit levels can be detected in other regions of the hippocampus (e.g., dentate gyrus) which are more resistant to ischemic injury [4]. Because AMPA receptors containing the GRIA2 subunit have low divalent cation permeability, the observed upregulation of GRIA2 is most probably protective in nature. Furthermore, changes in AMPA receptor subunit expression can also be observed following chronic administration of such drugs as

* Corresponding author. Fax: +65 67748056.
E-mail address: chongallen@yahoo.com (A. Chong).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.11.009

pyschotropics, pyschostimulants, antidepressants, and antipyschotic medications. Thus, we can clearly see here that AMPA receptor expression is dynamic.

Initially, we sought key transcription factor binding sites (TFBSs) within the GRIA promoters that are essential to the regulation of the expression of GRIA subunits. The identification of key transcriptional regulatory elements in the GRIA promoters has a greater global significance in that it can be applied in the design of novel gene-targeting constructs. For example, the identification of a neuronspecific glutamate receptor promoter element could possibly be used to deliver future experimental transgene and therapeutic agents to selected neurons in the brain.

Using a series of biocomputing procedures and statistical processes, we identified a combination of individual TFBSs and composite elements that are conserved and overrepresented within the promoters of the human, mouse, and rat GRIA genes. A composite element is a set of TFBSs found in combination and, usually, in close proximity to each other on the promoter, working synergistically to control the expression of a gene. An example is the IL-4-responsive element in the SOCS-1 promoter which contains three STAT6 and one Ets consensus binding sequences [5]. Ets-1 is confirmed to physically interact with STAT6 and IL-4 responsiveness was either partially or totally abolished following specific mutations. In the course of this work, we also identified genes which we believe are transcriptionally coregulated with GRIAs due to their shared promoter features with GRIA promoters. These coregulated genes would be expected to have an expression profile that correlates well with GRIAs. DNA microarrays represent a useful tool to unravel complex biological networks on a genomewide scale. The extensive use of microarray technology has resulted in large repositories of expression data being made available to the public [6,7]. Furthermore, recent studies have shown that if the expression of two or more genes is constantly related throughout many independent microarray datasets, the genes are highly likely to be functionally related [8,9]. Through the analysis of publicly available gene expression data, we identified 24 genes, that share not only a promoter profile but also an expression profile with GRIAs, which we believe are transcriptionally co-regulated with GRIAs.

Results and discussion

Promoter profiling the GRIA family

Conserved and overrepresented transcriptional elements within GRIA promoters. Just as it is believed that there are gene-specific and tissue-specific transcriptional elements, we believed that there might be a combination of transcriptional regulatory elements within the promoters of a gene family that uniquely controls their expression and that would be conserved in evolution across the promoters of this gene family in various species of organisms. To find the family-specific transcriptional elements of the AMPA receptor genes, we used the approach developed by Bajic *et al.* [10] on the promoters of these genes from the human, rat, and mouse. As there are only four human GRIA genes, this represents a very small dataset on which to work. Thus, we also used the promoters of four orthologous genes from mouse and one orthologous gene from rat. We used the promoters of these GRIA genes from three species to identify conserved and overrepresented transcriptional elements, while using some 10,000-odd promoter sequences as a background for comparison. This could be viewed as an advanced form of phylogenetic footprinting. However, we believe that if the gene family was sufficiently big, this technique could be applied to a single gene family of a single species.

Computational analysis on a single promoter sequence often throws up false positive errors; for example, Borges and Dingledine [11], using MatInspector, found two putative AP-1 sites within the GRIA1 promoter but electrophoretic mobility shift assays failed to show binding of c-Jun to these sites. However, our method of feature identification reduces these errors and provides a higher level of statistical significance because we are identifying not the individual TFBSs or composite elements on a single promoter sequence but the combination of regulatory elements that is characteristic to several promoter sequences in a gene family.

Based on our analysis, we selected the top-three-ranked "single," "pair," and "triplet" TFBS patterns that were overrepresented within **all** nine GRIA promoters that we studied here (The positions of the "singles," "pairs," and "triplets," with respect to the transcription start site [TSS], can be found in the Supplementary Data). We present the top three overrepresented "single," "pair," and "triplet" transcriptional element of the GRIA promoters in Tables 1a, 1b and 1c.

Various calculations can be used to demonstrate the statistical significance of the overrepresented transcriptional elements found in the target promoter set (9 GRIA promoters) (as contrasted against the background promoter sequence set of 10,741 human promoters). However, one should bear in mind that biological relevance of a pattern does not invariably mean that the pattern should be statistically significant. For example, the initiator element that is required for the initiation of transcription of many eukaryotic genes is a statistically insignificant pattern. For this reason, we opted not to select the patterns in the target promoter group based on any statistical parameter, except that the patterns have to be present in all members of the GRIA family that we examined. Nevertheless, for the sake of completeness, we provide in Tables 1a, 1b, and 1c two sets of statistics that reflect the statistical significance of these patterns: the overrepresentation index (ORI) and the p value of the patterns.

This combination of TFBSs and composite elements characterized the promoters of the GRIA family well. That is to say, this combination of 3 "singles," 3 "pairs," and 3 "triplets" could be used to distinguish the GRIAs' promoters from among 10,741 gene promoters, although with a small degree of what we initially thought were "false positive" errors. These "false positives" are 47 gene promoters from the pool of 10,741 promoters that were found to also have this same combination of "singles," "pairs," and "triplets." Although this promoter profile of 3 "singles," 3 "pairs," and 3 "triplets" is not entirely unique to the GRIA promoters, it is still significant as it is

represented in less than 0.5% of a comprehensive group of gene promoters. This led us to speculate that the reason that these 47 gene promoters share GRIAs' promoter profile is because they are coregulated/coexpressed with the GRIA genes and play an important role in GRIA expression or function. These 47 genes are listed and broadly categorized into 14 functional groups, ranging from cellular trafficking to transcriptional regulation in Supplementary Table 1.

Composite elements within GRIA promoters. From among the TFBSs and composite elements that we identified to be overrepresented on all nine AMPA receptor gene promoters, we found that there is prior evidence for the existence of the composite element pair, MZF1 and GATA (see Table 1b). Yu and colleagues [12] found that the solitary ERV-9 long terminal repeat located upstream of the HS5 site in the human beta-globin in erythroid K562 cells contained DNA motifs that bound the ubiquitous factor, NF-Y, and MZF1 and GATA-2. Through protein/protein interactions, NF-Y bound at the CCAAT motif and recruited MZF1 and GATA-2 and stabilized their binding to the neighboring GTGGGGA and GATA motifs.

Results suggest that angiotensin (AngII) activates STAT6 and STAT3 and these transcription factors are involved in the activation of the angiotensinogen (ANG) promoter via their recognition of the St-domain sequence [13]. In addition, a STAT3/Lyf-1/MZF1 composite element located in the promoter region from −238 to −144 of the mouse frizzled-related protein 4 (sFrp4) gene was found to be essential for the promoter activity of sFrp4 [14]. This suggests that STAT3 and MZF1 may interact with one another and thus leads us indirectly to believe that STAT6/MZF1/STAT3 may interact and bind to the composite element triplet that we identified on the GRIA promoters (Table 1c).

Cytokine-induced activation of the Fcgamma receptor I promoter required the DNA binding and the transactivation functions of both Stat1 and PU.1 [15]. In addition, an analysis of the human CD40 promoter indicates that the two gamma activated sequence sites at −521 and −483 and two Ets family member binding sites located at −553 and −447 are important for interferon (IFN)-gamma induction of CD40 transcription [16]. PU.1/Spi-B binds the distal (−553) while PU.1 binds the proximal (−447) Ets sites. How these transcription factors cooperate to switch on CD40 promoter activity is unclear but their close proximity might suggest a direct interaction of STAT1/PU.1/Spi-B. Further evidence of STAT1/PU.1 cooperativity can also be seen in IFN-gamma's induction of transcription of macrophage fgl2 gene [17]. Incidentally, it is believed that IFN-gamma treatment of spinal dorsal horn neurons causes a reduced expression of GluR1 [18]. It should also be noted that IFN-gamma can increase the expression of GKLF [19]. We therefore postulate that the GKLF/PU.1/STAT1 composite element that we identified (Table 1c) may play a role in the IFN-gamma-mediated decrease of GluR1 expression.

Analysis of genes that are coexpressed with GRIAs

To support our claim that the above 47 genes are coregulated/coexpressed with GRIAs, we used available expression data repositories and tools to find evidence to show that these genes and GRIAs have a statistically correlated coexpression pattern. We used publicly available resources to find genes that show a pattern of expression similar to that of the GRIA family of genes.

Gene sorter. The first resource is provided by the UCSC Genome Browser called Gene Sorter [20]. Mining through the various expression databases available on Gene Sorter, we

Table 1
Top-three-ranked overrepresented (a) "Singles," (b) "Pairs," and (c) "Triplets" found within GRIA promoters

(a)

	TFBS	Strand	ORI	*p* value
1.	CDP CR1	−ve	1.5499	0.02153
2.	Sp3	+ve	1.5727	0.01888
3.	Bach2	+ve	1.7398	0.00760

(b)

	5′→3′					
	TFBS	Strand	TFBS	Strand	ORI	*p* value
1.	GKLF	+ve	PU.1	+ve	2.5702	0.00022632
2.	MZF1	+ve	GATA-2	+ve	2.5796	0.00021901
3.	PU.1	+ve	GKLF	+ve	2.6380	0.00017899

The order in which the TFBS appears in each pair is given above from the 5′→3′ direction.

(c)

	5′→3′							
	TFBS	Strand	TFBS	Strand	TFBS	Strand	ORI	*p* value
1.	STAT6	−ve	MZF1	+ve	STAT3	−ve	4.3538	2.1958e-006
2.	ELF-1	+ve	STAT1	−ve	Pax-4	−ve	4.4384	1.8460e-006
3.	GKLF	+ve	PU.1	+ve	STAT1	−ve	5.5722	2.3734e-007

The order in which the TFBS appears in each triplet is given above from the 5′→3′ direction.

found 4793 genes that are closely coexpressed with GRIAs. Of the 47 genes that share GRIAs' promoter profile, we found that 16 are among this list of 4793 closely coexpressed genes identified by Gene Sorter (Table 2). In other words, 16 genes were identified to share not only a promoter profile but also a similar expression profile.

Gene expression omnibus (GEO). Next, we used National Center for Biotechnology Information's (NCBI) gene expression omnibus to perform a similar search for genes that have an expression profile correlated to that of the GRIAs [7]. Here, we found 7354 such genes (termed "profile neighbors" by GEO) expressed in brain, of which 10 are among the list of 47 genes that share GRIAs' promoter profile (Table 2). In other words, 10 genes were found to be closely coexpressed with GRIAs while also sharing GRIA's promoter profile. However, of these 10 genes, 6 were the same as those identified by Gene Sorter.

CLEO Database (based on data from the Stanford Microarray Database). With data provided by Ferdinando DiCunto from the CLEO database [21], we analysed the top 1% of genes with a coexpression pattern statistically correlated to that of the GRIAs from the human and mouse gene expression data of the Stanford Microarray Database [6]. Here, we found that of the 47 genes that share GRIAs' promoter profile, 10 are among the top 1%—6 genes (of 2202 human genes that were found closely coexpressed with GRIA) from the human gene expression data and 4 genes (of 1287 mouse genes that were found closely coexpressed with GRIA) from the mouse gene expression data (Table 2). In other words, 10 genes were found to be closely coexpressed with GRIAs while also sharing GRIA's promoter profile. However, of these 10, 6 were the same as those identified by either Gene Sorter and/or GEO.

There are approximately 20,000 confirmed protein-coding human genes (The International Human Genome Sequencing Consortium confirms the existence of 19,599 protein-coding genes). Examining the data for the human gene population, we calculated the *p* value for enrichment to be $4.046736e^{-002}$ (after conservative correction for multiplicity testing done with the Bonferroni method) for genes that are closely coexpressed with GRIAs while also sharing a similar promoter profile (See Supplementary Information for details). Furthermore, support for the coexpression with GRIAs is confirmed by two or more of the above-mentioned analyses for 10 genes (that is, our analyses of the human and mouse gene expression data, NCBI's GEO, and/or the human Gene Sorter program) (Table 2).

For the UCSC Gene Sorter results, the similarity in expression of two genes is calculated by a weighted sum of differences in log expression ratio values, whereas, for our analysis of the data from the Stanford Microarray Database and for GEO Profiles' precalculated profile neighbors, a calculation of the Pearson's correlation coefficient was made to acertain the closeness in expression of two genes. If one were to use the UCSC results as a point of reference, then by two different methods of expression profiling we can find at least eight different genes that are not only confirmed to be coexpressed with GRIA but also share a similar promoter profile. These eight genes (BAI2, IL16, KLK6, CLSTN3, FAM107A, KLHL18, BEX1, DSN1) therefore represent very strong candidates for further laboratory studies.

Table 2
Genes that are closely coexpressed with GRIAs as determined by analyses of independent gene expression datasets

	Human Gene ID (Mouse Gene ID given in brackets)	Gene Symbol	Stanford Microarray Database (Human)	Stanford Microarray Database (Mouse)	UCSC Human Gene Sorter	NCBI GEO
1	576	BAI2	✓		✓	✓
2	911	CD1C	✓			
3	2323 (14256)	FLT3LG		✓		
4	3603	IL16			✓	✓
5	4793	NFKBIB	✓			
6	5279	PIGC	✓			✓
7	5653 (19144)	KLK6		✓	✓	
8	5865	RAB3B			✓	
9	8674	VAMP4				✓
10	9746	CLSTN3			✓	✓
11	11131	CAPN11			✓	
12	11170	FAM107A	✓		✓	✓
13	23276	KLHL18			✓	✓
14	25852	ARMC8				✓
15	27120 (50722)	DKKL1		✓		✓
16	54093	SETD4			✓	
17	54897	CASZ1			✓	
18	55244 (67473)	FLJ10847		✓		
19	55859	BEX1			✓	✓
20	56999	ADAMTS9			✓	
21	64225	ARL6IP2			✓	
22	64577	ALDH8A1			✓	
23	79980	DSN1	✓		✓	
24	80227	WDR71			✓	

Coexpression data support hypothesis of coregulated genes

By coregulation, we mean that the transcriptional regulation of a particular gene is closely linked to the transcription of GRIA genes due to their shared promoter characteristics. Coregulated genes are believed to be functionally related and play a significant role in aiding the function and expression of their partner. While coregulation of two genes may result in the coexpression of these two genes in the same tissue, coexpression of two genes in the same tissue does not necessarily mean that they are transcriptionally coregulated since their coexpression may be coincidental. However, we should not also expect two transcriptionally coregulated genes to always be coexpressed since their transcriptional coregulation may take place only under the right physiological conditions; transcription, after all, is affected by various factors, such as, tissue specificity. Recent studies have shown that if the expression of two or more genes is constantly related throughout many independent microarray datasets, the genes display a significant degree of functional similarity [8,9]. In this context, it is interesting to note that 10 genes were repeatedly shown to be coexpressed with GRIAs by our analysis of the human and mouse data from the Stanford Microarray Database and by the Gene Sorter tool and/or NCBI's GEO. Particularly interesting is the close coexpression of KLK6 with GRIA that was found in human (by Gene Sorter) and in mouse (Stanford Microarray Database) which strongly suggests a functional relationship between these two genes. Phylogenetic

conservation has been proposed as a very strong criterion for identifying functionally relevant coexpression links between genes [22]. Conservation implies that the coexpression of the gene pairs confers a selective advantage and, therefore, these genes are most likely functionally related.

Here, we have coupled promoter profiling with the analysis of coexpression data from various microarray experiments. Together, they lend support to one another in suggesting that at least 24 genes that we have identified are functionally coregulated (see Table 2) since the chance of 2 genes sharing similar promoter characteristics and expression patterns goes beyond mere coincidence (please refer to the online Supplementary Information for a calculation of the p value for finding genes that both have a highly correlated coexpression with the GRIAs and share the same set of promoter elements as those that we selected). Among these 24 genes which we believe are coregulated with GRIAs are Rab3B and calsyntenin 3 (CLSTN3).

Rab3B belongs to the Rab3 family of small GTP-binding proteins which include Rab3A, Rab3C, and Rab3D that function in regulated exocytosis in various secretory cells [23]. Recruitment of AMPA receptors to the synapse is driven by calmodulin-dependent protein kinase II (CAMKII) [24]. Incidentally, Rab3B binds and interacts with Ca^{2+}-calmodulin (CaM) in a calcium-dependent manner and was shown to be involved in Ca^{2+}-dependent exocytosis in rat anterior pituitary cells [25,26]. Immunocytochemistry studies confirm the presence of GRIA1- and GRIA2/3-positive cells in the anterior and intermediate lobes of the pituitary [27]. Also, it is thought that the delivery of AMPA receptors to the synaptic membrane occurs through an exocytic pathway [28]. The localization of Rab3B on synaptic vesicles and its demonstrated interaction with Ca^{2+}-CaM in exocytosis, coupled with its matching expression profile with GRIAs' in the brain, provide a strong indication that Rab3B might be enlisted to help in the surface expression of AMPA glutamate receptors.

CLSTN3 belongs to a recently discovered family of novel postsynaptic membrane proteins and was identified as a target protein of extracellular proteases [29]. Synaptic plasticity is accompanied by structural changes in and around the synapse and these structural reorganizations are due to the action of extracellular proteases, such as tissue plasminogen activator and neuropsin [30–32]. The exact function of calsyntenins is unknown; however, immunoelectron microscopy reveals that all three calsyntenins are located in the postsynaptic membrane of asymmetrical (excitatory) synapses [29]. There is ample evidence showing that excitatory AMPA glutamate receptors are also found on postsynaptic dendritic spines of asymmetrical synapses [33–36]. In addition, Hintsch and co-workers [29] demonstrated that several clearly discernible populations of glutamatergic neurons, such as those in neocortical layer 5 and the hippocampal CA1-CA3 regions, expressed high levels of CLSTN3. Because CLSTN3 is a target of extracellular proteases that play a central role in synaptic plasticity, we believe that CLSTN3 is coexpressed/coregulated in neurons undergoing AMPA-receptor-mediated synaptic plasticity to aid in structural reorganization of the synapse.

One pertinent question that arises is what if microarray expression data do not show a coexpression of a particular gene (say, Gene X) with GRIA? Can we definitively say that this is a false positive prediction? The answer is, of course, no. The circumstances under which Gene X would be coregulated with GRIA may not be present during the microarray experiment (for example, perhaps the cells need to be treated with tumor necrosis factor-alpha for us to see a coregulated expression of Gene X and GRIA).

In further support of our contention that the 47 genes are coregulated with GRIAs, a search for the expression profiles of all 47 genes in UCSC's Gene Sorter and GEO confirms that all 47 genes are expressed in the brain (fetal and/or adult) to varying degrees. This is important since AMPA receptors are found predominantly in the nervous system.

Conclusion

Our initial aim was to find a family-specific set of transcriptional elements that were significantly conserved across the GRIA promoters. We found, instead, a set of transcriptional elements that was shared by GRIAs and their functionally related coregulated genes. In other words, we found a function-specific set of transcriptional elements. Promoter profiling helped us identify unique promoter features common among functionally related genes which control their transcription and thus, in this process, we were able to identify what we believed are genes coregulated with GRIAs. Experimental validation of their coregulated expression was obtained from the wealth of freely available microarray expression data. Recent studies have shown that two genes are functionally related if their expression is consistently related in various independent microarray datasets [8,9]. Even so, as we can see, for example from GEO, an analysis of various expression datasets can yield thousands of so-called profile neighbors. Therefore, on their own, neither of these two methods can definitively identify functional groups of coregulated genes, however, by combining these two methods (promoter and expression profiling), we provide strong justification to suggest that at least 24 novel genes are functionally related to and transcriptionally coregulated with the GRIA family. This conclusion is based on three characteristics: a common promoter profile, a common expression profile, and high statistical significance. More work will need to be done in the laboratory to fully elucidate the role of these genes in AMPA receptor physiology; however, the results that we present here give a solid foundation on which to base future investigations/search for interacting partners of AMPA receptors.

A different approach to take would have been to reverse the process followed here, that is to first identify a small group of genes tightly coexpressed with GRIA and then proceed with promoter profiling. However, the argument against such an approach is that, by first clustering genes based on expression profiles, you obtain clusters that contain not only one group of coregulated genes but generally many different groups of coregulated genes placed in the same coexpression cluster. Because of this, such an approach produces clusters that

dominantly contain a group of genes coregulated in the same manner only in exceptional cases.

There are a number of published methods that relate to combined analysis of motifs in promoters and their links to expression [37–42]. However, none of these methods allows us to use efficiently the whole set of TRANSFAC matrix models for the human promoter set that we had (10741) and to generate automatically combinations of patterns for "pairs" and for "triplet" as our in-house developed tool did. We, however, do not exclude the possibility that results similar to those presented in this paper could be obtained by some of the other available tools [37–42].

Methods

Promoter sequence collection

Promoter sequences (10,741) of human genes covering the region −1500 to +1000 (with respect to the transcription start site) were collected by the FIE2 program [43]. The release of the human genomic sequences at the time of collection was NCBI Build 31. The extraction of the human GRIA promoters was done similarly. For the mouse GRIA promoter sequences, the alignments of the individual mouse GRIA genes against the then available mouse genomic sequences from the Mouse Genome Sequencing Consortium (MGSCv3) were obtained from NCBI's LocusLink [44]. The 5′ ends of the extracted mouse GRIA sequences were passed through the Dragon Promoter Finder [45] to predict the TSS for these genes. At the start of this study, work on the rat genomic sequence was at its threshold and so only the rat GRIA1 promoter sequence (AF302117) [11] was used. Five TSSs were previously identified for the rat GRIA1 gene. For this study, we chose the 5′-most TSS as a point of reference.

Comparison of target promoters against background promoters

GRIA promoters (target promoters) were compared with the 10,741 promoters (background promoters) collected by FIE2 to determine the overrepresented transcriptional elements within GRIA promoters. To do this, we first mapped all TFBSs from TRANSFAC Professional database ver. 6.2 [46] to all promoter sequences. This mapping was carried out using the MATCH program [47] with the "minsum" setting. This parameter setting allows for the minimized sum of false positive and false negative predictions of TFBSs. Once that was done, we compared densities of individual TFBSs in the target and background promoters and calculated the overrepresentation for each TFBS, that is determining how much more dense a particular TFBS is in the GRIA promoters with respect to the background promoters following the procedure that we previously described [10]. These individual TFBSs were termed "singles." Similarly, we repeated the analysis for all combinations of paired TFBSs where the two TFBSs were no more than 50 nt apart (which we termed "pairs"). The same analysis was also carried out for combinations of three TFBSs, with maximal mutual distance of neighboring TFBSs not greater than 50 nt (which we termed "triplets").

Identification of closely coexpressed genes

An analysis of human and mouse gene expression data from the Stanford Microarray Database [6] was performed to obtain the top 1% of genes which are closely coexpressed with GRIAs. This was done by the method previously described by Pellegrino *et al.* [21]. We also performed a search for genes with similar expression patterns using the Human Gene Sorter tool [20]. The Gene Sorter calculates and displays genes by their similarity in expression to a selected gene. The similarity is calculated as a weighted sum of differences in log expression ratio values using the expression data from such sources as the GNF Gene Expression Atlas 2 [48]. In addition, we used the NCBI GEO "profile neighbor" function to obtain a list of genes that are closely related in expression to the GRIAs, based on data deposited in GEO [7].

P value calculation

We used the right-sided Fisher's exact test based on hypergeometric distribution. Where the p values are corrected by the Bonferroni method for multiplicity testing, this is explicitly indicated.

Acknowledgments

We extend our warmest thanks to Tanya Barrett (NCBI GEO) and Jim Kent (UCSC) for their time and assistance. In addition, we are deeply indebted to Professor Ferdinando DiCunto for providing data from the CLEO database and for his valuable comments and discussion on this work. We are also very grateful to our colleagues, Professors Limsoon Wong and Cathal Seoighe, for their invaluable comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2006.11.009.

References

[1] F.E. Jensen, The role of glutamate receptor maturation in perinatal seizures and brain injury, Int. J. Dev. Neurosci. 20 (2002) 339–347.
[2] R.A. Nicoll, Expression mechanisms underlying long-term potentiation: a postsynaptic view, Philos. Trans. R. Soc. Lond., B Biol. Sci. 358 (2003) 721–726.
[3] J.A. Gorter, et al., Global ischemia induces downregulation of Glur2 mRNA and increases AMPA receptor-mediated Ca2+ influx in hippocampal CA1 neurons of gerbil, J. Neurosci. 17 (1997) 6179–6188.
[4] H. Pollard, A. Heron, J. Moreau, Y. Ben-Ari, M. Khrestchatisky, Alterations of the GluR-B AMPA receptor subunit flip/flop expression in kainate-induced epilepsy and ischemia, Neuroscience 57 (1993) 545–554.
[5] J. Travagli, M. Letourneur, J. Bertoglio, J. Pierre, STAT6 and Ets-1 form a stable complex that modulates Socs-1 expression by interleukin-4 in keratinocytes, J. Biol. Chem. 279 (2004) 35183–35192.
[6] J. Gollub, et al., The Stanford Microarray Database: data access and quality assessment tools, Nucleic Acids Res. 31 (2003) 94–96.
[7] T. Barrett, et al., NCBI GEO: mining millions of expression profiles—Database and tools, Nucleic Acids Res. 33 (2005) D562–D566.
[8] H.K. Lee, A.K. Hsu, J. Sajdak, J. Qin, P. Pavlidis, Coexpression analysis of human genes across many microarray data sets, Genome Res. 14 (2004) 1085–1094.
[9] M.N. Price, E. Rieffel, Finding coexpressed genes in counts-based data: an improved measure with validation experiments, Bioinformatics 20 (2004) 945–952.
[10] V.B. Bajic, V. Choudhary, C.K. Hock, Content analysis of the core promoter region of human genes, In Silico Biol. 4 (2004) 109–125.
[11] K. Borges, R. Dingledine, Functional organization of the GluR1 glutamate receptor promoter, J. Biol. Chem. 276 (2001) 25929–25938.
[12] X. Yu, et al., The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2, J. Biol. Chem. 280 (2005) 35184–35194.
[13] E. Mascareno, M. Dhar, M.A. Siddiqui, Signal transduction and activator of transcription (STAT) protein-dependent activation of angiotensinogen promoter: a cellular signal for hypertrophy in cardiac muscle, Proc. Natl. Acad. Sci. USA 95 (1998) 5590–5594.
[14] V.K. Wong, J.W. Yam, W.L. Hsiao, Cloning and characterization of the promoter region of the mouse frizzled-related protein 4 gene, Biol. Chem. 384 (2003) 1147–1154.
[15] S. Aittomaki, J. Yang, E.W. Scott, M.C. Simon, O. Silvennoinen, Molecular basis of Stat1 and PU.1 cooperation in cytokine-induced

Fcgamma receptor I promoter activation, Int. Immunol. 16 (2004) 265–274.
[16] V.T. Nguyen, E.N. Benveniste, Involvement of STAT-1 and ets family members in interferon-gamma induction of CD40 transcription in microglia/macrophages, J. Biol. Chem. 275 (2000) 23674–23684.
[17] M. Liu, et al., Cytokine-Induced Hepatic Apoptosis Is Dependent on FGL2/Fibroleukin: The Role of Sp1/Sp3 and STAT1/PU.1 Composite cis Elements, J. Immunol. 176 (2006) 7028–7038.
[18] K.S. Vikman, R.H. Hill, E. Backstrom, B. Robertson, K. Kristensson, Interferon-gamma induces characteristics of central sensitization in spinal dorsal horn neurons in vitro, Pain 106 (2003) 241–251.
[19] Z.Y. Chen, J.L. Shie, C.C. Tseng, STAT1 is required for IFN-gamma-mediated gut-enriched Kruppel-like factor expression, Exp. Cell Res. 281 (2002) 19–27.
[20] W.J. Kent, et al., Exploring relationships and mining data with the UCSC Gene Sorter, Genome Res. 15 (2005) 737–741.
[21] M. Pellegrino, P. Provero, L. Silengo, F. Di Cunto, CLOE: identification of putative functional relationships among genes by comparison of expression profiles between two species, BMC Bioinformatics 5 (2004) 179.
[22] J.M. Stuart, E. Segal, D. Koller, S.K. Kim, A gene-coexpression network for global discovery of conserved genetic modules, Science 302 (2003) 249–255.
[23] O.M. Schluter, M. Khvotchev, R. Jahn, T.C. Sudhof, Localization versus function of Rab3 proteins. Evidence for a common regulatory role in controlling fusion, J. Biol. Chem. 277 (2002) 40919–40929.
[24] J.A. Esteban, AMPA receptor trafficking: a road map for synaptic plasticity, Mol. Interv. 3 (2003) 375–385.
[25] P.M. Lledo, P. Vernier, J.D. Vincent, W.T. Mason, R. Zorec, Inhibition of Rab3B expression attenuates Ca(2+)-dependent exocytosis in rat anterior pituitary cells, Nature 364 (1993) 540–544.
[26] R.S. Sidhu, R.P. Bhullar, Rab3B in human platelet is membrane bound and interacts with Ca(2+)/calmodulin, Biochem. Biophys. Res. Commun. 289 (2001) 1039–1043.
[27] H. Kiyama, K. Sato, M. Tohyama, Characteristic localization of non-NMDA type glutamate receptor subunits in the rat pituitary gland, Brain Res. Mol. Brain Res. 19 (1993) 262–268.
[28] G. Broutman, M. Baudry, Involvement of the secretory pathway for AMPA receptors in NMDA-induced potentiation in hippocampus, J. Neurosci. 21 (2001) 27–34.
[29] G. Hintsch, et al., The calsyntenins–a family of postsynaptic membrane proteins with distinct neuronal expression patterns, Mol. Cell. Neurosci. 21 (2002) 393–409.
[30] D. Baranes, et al., Tissue plasminogen activator contributes to the late phase of LTP and to synaptic growth in the hippocampal mossy fiber pathway, Neuron 21 (1998) 813–825.
[31] A. Hirata, et al., Abnormalities of synapses and neurons in the hippocampus of neuropsin-deficient mice, Mol. Cell. Neurosci. 17 (2001) 600–610.
[32] T. Oka, et al., Extracellular serine protease neuropsin (KLK8) modulates neurite outgrowth and fasciculation of mouse hippocampal neurons in culture, Neurosci. Lett. 321 (2002) 141–144.
[33] A. Baude, Z. Nusser, E. Molnar, R.A. McIlhinney, P. Somogyi, High-resolution immunogold localization of AMPA type glutamate receptor subunits at synaptic and non-synaptic sites in rat hippocampus, Neuroscience 69 (1995) 1031–1055.
[34] V. Bernard, P. Somogyi, J.P. Bolam, Cellular, subcellular, and subsynaptic distribution of AMPA-type glutamate receptor subunits in the neostriatum of the rat, J. Neurosci. 17 (1997) 819–833.
[35] N.P. Clarke, J.P. Bolam, Distribution of glutamate receptor subunits at neurochemically characterized synapses in the entopeduncular nucleus and subthalamic nucleus of the rat, J. Comp. Neurol. 397 (1998) 403–420.
[36] J.P. Kessler, A. Baude, Distribution of AMPA receptor subunits GluR1-4 in the dorsal vagal complex of the rat: a light and electron microscope immunocytochemical study, Synapse 34 (1999) 55–67.
[37] R.S. Kim, H. Ji, W.H. Wong, An improved distance measure between the expression profiles linking co-expression and co-regulation in mouse, BMC Bioinformatics 7 (2006) 44.
[38] J. Ruan, W. Zhang, A bi-dimensional regression tree approach to the modeling of gene expression regulation, Bioinformatics 22 (2006) 332–340.
[39] L.H. Lin, H.C. Lee, W.H. Li, B.S. Chen, Dynamic modeling of cis-regulatory circuits and gene expression prediction via cross-gene identification, BMC Bioinformatics 6 (2005) 258.
[40] A.G. Jegga, et al., CisMols Analyzer: identification of compositionally similar cis-element clusters in ortholog conserved regions of coordinately expressed genes, Nucleic Acids Res. 33 (2005) W408–W411.
[41] M.E. Hudson, P.H. Quail, Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data, Plant Physiol. 133 (2003) 1605–1616.
[42] Y. Pilpel, P. Sudarsanam, G.M. Church, Identifying regulatory networks by combinatorial analysis of promoter elements, Nat. Genet. 29 (2001) 153–159.
[43] A. Chong, G. Zhang, V.B. Bajic, FIE2: A program for the extraction of genomic DNA sequences around the start and translation initiation site of human genes, Nucleic Acids Res. 31 (2003) 3546–3553.
[44] K.D. Pruitt, D.R. Maglott, RefSeq and LocusLink: NCBI gene-centered resources, Nucleic Acids Res. 29 (2001) 137–140.
[45] V.B. Bajic, et al., Dragon Promoter Finder: recognition of vertebrate RNA polymerase II promoters, Bioinformatics 18 (2002) 198–199.
[46] V. Matys, et al., TRANSFAC: transcriptional regulation, from patterns to profiles, Nucleic Acids Res. 31 (2003) 374–378.
[47] A.E. Kel, et al., MATCH: A tool for searching transcription factor binding sites in DNA sequences, Nucleic Acids Res. 31 (2003) 3576–3579.
[48] A.I. Su, et al., A gene atlas of the mouse and human protein-encoding transcriptomes, Proc. Natl. Acad. Sci. USA 101 (2004) 6062–6067.